Chapter 2: Records and Patient Sets

In the SEER*DMS database, patient data are stored in "records" and "patient sets." Records contain the source data submitted to the registry, including medical data submitted by reporting facilities and follow-up information provided by non-medical organizations. Each record contains a single report of data, for example, a NAACCR record contains data from a single abstract. Patient sets contain all data associated with a particular patient. Each patient set contains the data consolidated from the patient's source records and linkage information to enable access to the source records. Patient set data fields contain the consolidated and/or summarized values of the linked source records, and values that were derived or computed by SEER*DMS polishers. Original data values are maintained in record data fields (and in an archived copy of the imported data file).

In this chapter, you'll learn about

- Records
  - Linked and Unlinked Records
- Patient Sets
  - Cancer/Tumor/Case (CTC)
  - Summary Treatment (Summary TX)
  - Treatment Procedures (TX and TXr)
  - Facility Admission
  - SEER Course
  - Incoming and Linked Records
  - Auto-building Patient Sets and CTCs
- Audit Logs
- Finding Patient Data in SEER*DMS
- SEER*DMS Database and Data Warehouse

Records

In the SEER*DMS database, the record table contains the source data submitted to the registry. Records are created in the database when data are loaded from files or entered manually (as described in *Chapter 5: Importing Data Files* and *Chapter 6: Data Entry*).

Each record travels through the workflow, triggering the automated and manual "tasks" that must be performed to process the data. The path of a record through the workflow is determined by the record's type. The specific workflow path of each record is documented in diagrams provided in the *Workflow Diagrams by Record Type* section of the SEER*DMS Web Portal (https://seer.cancer.gov/seerdms/portal. The following is a list of all record types supported by SEER*DMS; some record types may not be available in your registry's configuration.

- **Health Record** – An initial abstract record submitted by a medical facility in a registry-defined layout. Health records move through the main portion of the workflow which involves editing, screening, matching, and consolidating tasks. If no match is found for a reportable health record, a new patient set will be created.
- **NAACCR Abstract** – An initial abstract record submitted by a medical facility in the NAACCR record layout or Iowa Transmit record layout (a multi-line format with one line in the NAACCR format followed by optional lines for expanded text fields). NAACCR Abstract records move through the main portion of the workflow which involves editing, screening, matching, and consolidating tasks. If no match is found for a reportable NAACCR Abstract record, a new patient set will be created.

- **NAACCR Modified Record** – An abstract record that was modified and re-submitted to the registry. This record is identical in format to the NAACCR record (the record type field is coded with an 'M').
- **NAACCR Update** – The NAACCR short format record used to submit field-specific corrections to data previously submitted to the registry. SEER*DMS matches the record to existing patient set data but does not specifically attempt to match the update record to the original NAACCR record. If SEER*DMS identifies possible matches but cannot identify a definite match, a manual matching task will be created. If SEER*DMS identifies a patient set that is an exact match, SEER*DMS attempts to auto-consolidate the data. A manual Consolidate FUP task will be created if the system cannot complete the auto-consolidation.
- **Short Health Record** – A SEER*DMS short health record is a partial or incomplete abstract. A partial abstract may include some patient demographic information, complete or partial histopathology information, treatment information, or other information not sufficient to complete a full case abstract. Typically, these records provide additional information regarding follow-up, treatment, and admissions related to a previously abstracted case. In rare situations, a short health record may contain the only data related to a case and may be used to build a new CTC.
- **Follow-up Only** – Data reported by a facility indicating a history of cancer. These data may provide follow-up information for existing patient sets. In the workflow, Follow-up Only records move through the automatic edit checks and the auto-matching task. If no match is found, they are held to be re-matched at a later time.
- **Casefinding** – Data obtained from medical reports which identify new potential cases. These records trigger the creation of Abstract Facility Leads. If an abstract is never obtained, a new CTC or patient set can be built from a casefinding record.
- **HL7 E-Path** – Electronically transmitted pathology reports in the NAACCR HL7 Version 2.3.1 E-Path Standard. These records trigger the creation of Abstract Facility Leads. If an abstract is never obtained, a new CTC or patient set can be built from a casefinding record.
- **Follow-up Transmit** – A short record containing follow-up information (date, contacts, current address, etc) for patients previously report by the facility.
- **Death Certificate** – A data record containing death certificate data items. These data are used to update follow-up variables for existing patients. If a death certificate indicates a new reportable cancer, an Abstract Facility Lead is created. If an abstract can not be obtained, death clearance follow-back procedures are performed.
- **Death Notice** – Contains information published in public death notices. These records are used to update follow-up variables for existing patients. In the workflow, Death Notice records move through the edits, screening, and matching task. If no match is found, they are held to be re-matched at a later time.
- **Health Index Record** - Data from hospital listings such as diagnostic index, discharge logs, or surgery logs. These data are used for quality assurance related to casefinding and may provide follow-up and treatment data that are consolidated into existing patient sets.
- **Supplemental** – Follow-up information received from non-medical organizations. These records are matched to the database and used in passive follow-up procedures.
- **ME Cohort** – Data from the Hawai`i component of the US Public Health Service's Multi-ethnic Cohort (MEC) Study which are linked annually to the Hawai`i Tumor Registry data.

Record type is a general classification that can be assigned to various kinds of data provided in multiple file formats. For example, the Supplemental record type is used for department of motor vehicle, CMS/HCFA, and voter registration data. Health Records may come from multiple sources (including your registry's abstracting tool) and in different file formats. Specific file layouts are defined in registry-specific configuration files. For information about your registry's configuration, please refer to your registry's copy of the *SEER*DMS Technical Reference: Registry-specific Information*.

In SEER*DMS, records provide a mechanism for storing and accessing the original data values submitted to the registry. Record data do not have to be error-free; records may have null data fields or fields with errors. Problems with the data are generally resolved in the patient set data fields at the time of consolidation. This is a new concept for most cancer registry data editors. In general, it is recommended that registry staff do not change record data fields except to correct errors that prevent the record from being processed in SEER*DMS (see *Chapter 7: Resolving Record Errors* for more information). All changes are documented in the record's audit log which includes the original data value for each field that was modified. In addition, the original data files are archived and can be accessed by your registry's system administrators.

Linked and Unlinked Records

A "link" is an association between a source record and a patient set. The link is established after the record has been matched against the database. Based on the outcome of the matching process, the incoming record will either be linked to an existing patient set during consolidation or linked to a newly created patient set. These processes are briefly described below. Complete details are provided in your registry's workflow diagrams (see the Documentation section of the SEER*DMS Web Portal).

- If the incoming record matches an existing patient set, the record is linked to the patient set and the data from the record are consolidated with the data in the existing patient set. This involves a combination of manual and automated processes that are determined by record type, reportability, and registry-defined workflow algorithms.
    - For example, an abstract record for a reportable cancer and its matching patient set are forwarded to a manual Consolidate task. In this task, a registrar uses the tools in the Patient Set Editor to complete the consolidation by linking the incoming record and incorporating its data into the summarized patient set data fields.
    - On the other hand, a Supplemental record is linked and consolidated to a matching patient set in an automated task as part of the passive follow-up process. A manual Consolidate FUP task is only created if discrepancies between the record and patient set prevent auto-consolidation. In most registry configurations, records that follow a passive follow-up route through the workflow include Supplemental, Death Notice, non-reportable Death Certificates, and Follow-up Only records.
- If an abstract record contains reportable data for a new patient, a new patient set is auto-built using the record data and the record is linked to the newly created patient set. NAACCR Abstract, and Health Records contain abstracted medical data and are used to auto-build patient sets.
- "Unlinked" records are records that are not consolidated into a patient set, including:
    - Records that have entered the system but are in the initial tasks of the workflow and have yet to be matched.
    - Records that do not match an existing patient set and do not contain enough data to create a patient set. In most registry configurations, only abstract records are used to create patient sets as they enter the workflow. A patient set can be manually created from unlinked, reportable Casefinding, Death Certificate, or Short Health records using the record editor or by executing one of the build system tasks (see *Chapter 28: System Administration*).

The Patient Set Editor provides access to all linked records, enabling users to review the raw data that contributed to the summarized and consolidated Patient Set.

Patient Sets

A "patient set" is a packet of data in the SEER*DMS database which includes all data associated with a particular patient. These data include patient demographics, information on all reportable cancers, admissions information, diagnostic procedures, treatment information and text documentation. A "patient set" also includes direct links to the original records that contributed to the patient set data.

Cancer/Tumor/Case (CTC)

Cancer/Tumor/Case (CTC) refers to an instance of the disease of interest. Generally speaking, a Cancer/Tumor/Case is a neoplasm with topography, histology, and behavior codes that meet the SEER, local, or special study case reporting guidelines. While this is usually a cancer or a tumor, some code combinations are not truly tumors and others are arguably not cancer. Case is sometimes used interchangeably with "tumor" or "cancer" but case can also be used to refer to a person in the database.

In the documentation, we sometimes use “tumor” instead of CTC, usually within a standard data item name. We may also use “cancer” or “tumor” or “case” to refer to an instance of a disease that has not yet been registered in the database.

In the SEER*DMS database, CTC is the name of a data entity within a patient set. All diagnostic, staging, and treatment data for a reportable neoplasm are stored in a CTC. The CTC data include summarized treatment information, as well as data regarding each admission and treatment for that particular CTC. A patient set may include multiple CTCs, one for each primary cancer diagnosed for the patient.

Summary Treatment (Summary TX)

Summary treatment fields are stored in the CTC table of the database and displayed on the Summary TX page of the patient set editor. These fields contain the consolidated treatment values and associated dates for the treatments related to the CTC. SEER*DMS applies an algorithm to arrive at a suggested summary treatment value for each treatment modality (i.e., surgery, chemotherapy, radiation). A summarized date of therapy is also provided. The suggested summary value may or may not be the best value when combinations of treatment are being consolidated. For example; two single agent chemotherapy entries may be consolidated to either a single agent chemotherapy summary treatment or multi-agent chemotherapy.

Visual editing and confirmation is required to ensure the best summary treatment values are entered into the consolidated patient set data. When a case is consolidated, all treatments (TX) for a given CTC should be carefully reviewed. A review flag must also be set to confirm that all of the treatment data (including dates) have been reviewed and confirmed for best values.

Treatment Procedures (TX and TXr)

CTC treatment data are stored in the *treatment_procedure* table of the database and displayed on TX and TXr pages of the patient set editor. The Facility ID number associated with the place of treatment is listed next to the page link in the patient navigation box. For example, TX (2344) indicates that this treatment was reported by and administered at facility FAC-2344.

TXr is typically used to identify any treatment that was reported by one facility, but given at a different facility. TXr may also be used for registry-specific designations, for example, it is also used to identify treatment procedures with Previous Flag = Yes in SEER*DMS MDCSS.

If the treatment facility is known, the facility number of both the reporting facility and the treatment facility will appear following the TXr. For example, TXr (6431~2344) indicates that

according to a cancer report submitted by facility FAC-2344, this treatment was given at facility FAC-6431. If 9999 appears instead of a valid facility number, it indicates that the facility is unknown. For example, TXr (9999~2344) indicates that according to a record submitted by facility FAC-2344, this treatment was given at an unknown facility. If the reporting facility field is blank for TXr data, ???? will be displayed in the navigation (this is not possible if the definition of TXr is limited to "a treatment reported by one facility, but given at another").

Facility Admission

The Facility Admission data catalogues information gathered from each patient encounter at a given facility. If a patient has multiple admissions or encounters at the same facility, the facility may provide admission data for each encounter. These data are stored in the *facility_admission* tables of the database.

SEER Course

SEER requires that all treatment administered during the first course of therapy be included for each reported case. See the *SEER Program Coding and Staging Manual* for definitions of first course of therapy. Some registries also collect information on subsequent therapy. This could be treatment for recurrence of cancer or for disease progression. SEER*DMS allows the registry to identify the treatment course and will distinguish between treatment courses when deriving suggested best values for treatment. The default value for treatment course is first course of therapy.

Incoming and Linked Records

When a patient set is viewed during a consolidation task, records are listed as either "incoming" or "linked". An incoming record is a record which is a tentative match to the patient set. The record is currently being processed in the workflow. It was selected during a Match-Consolidate task for consolidation with the patient set data. If the match is confirmed during the consolidation task, the record will be linked to the patient set and the data will be consolidated. For more information please refer to *Chapter 12: Consolidating Data*.

Auto-building Patient Sets and CTCs

An entire patient set or components of an existing patient set must be created when new reportable data are incorporated into the database. For example, a new patient set must be created when it is determined that data for a reportable cancer are related to a new patient. A new CTC must be created within an existing patient set when it is determined that data matching an existing patient represents a new, reportable tumor.

To reduce the work involved in creating a new patient set, SEER*DMS auto-builds patient set components based on record data. The transfer of record data to the patient set is automated, when possible. The values of the appropriate data fields are auto-filled, as determined by registry-specific algorithms which are specified for each record type. For some fields, the value is copied directly from a record field to a corresponding patient set field. The values of other patient set data fields are imputed based on multiple fields in the record, or by recoding the value of a single field in the record. The field mappings and recoding algorithms are documented in the *SEER*DMS Technical Reference: Registry-Specific Information*.

These are the circumstances under which auto-building occurs:

- A reportable abstract record does not match an existing patient set. The system auto-builds a new patient set based on that record. A single CTC is created within the patient set. Within that CTC, the number of auto-built admissions and treatments will vary based on the record type and the amount of data available on the record.

- A reportable abstract record matches a patient already in the database, but describes a cancer for which that patient has no CTC. During the Consolidate task, a CTC is auto-built when the record is linked to a "New CTC" in the patient set. Within that CTC, the number of auto-built admissions and treatments will vary based on the record type and the amount of data available on the record.
- A reportable record matches an existing patient and CTC, but describes a new facility admission or treatment relevant to this CTC. In this case, the system attempts to auto-build new facility and treatment data fields for that CTC.

SEER*DMS auto-builds patient set data fields using data from a single record. If multiple records for a new patient are consolidated, the patient set data fields are auto-built from one of the records when the Consolidate task is started. SEER*DMS selects an abstract record (NAACCR Abstract, NAACCR Modified, or Health Record) to use for auto-building. If there are two or more abstract records, the abstract record that was first loaded into SEER*DMS is used. The other incoming records are then consolidated with the new patient set.

Audit Logs

Whenever a record or patient set data field is changed by a user or the system, an entry is made in an audit log. A separate audit log is maintained for each record and patient set in the system. The information in the audit log includes the user or process which modified the data, comments related to the change, the date and time of the modification, and the original and modified value of each data field that was changed. A user may provide separate comments for the individual data items or one comment for all data items changed within a single session.

When the audit log is displayed in the record or patient set editor, the changes are grouped by events. Events include workflow tasks, system tasks, and other processes that modify patient data. The audit log events in SEER*DMS are:

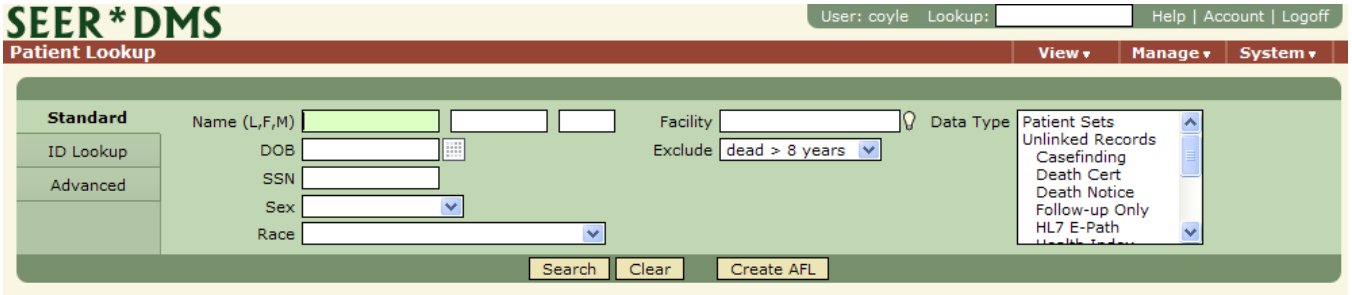
- **Ad-hoc Editing** – The change was made by a user who edited the data while accessing the data via the patient lookup, a worklist task was not involved.
- **AFL Update** – Changes made to an AFL were incorporated into the record that triggered the AFL. At this time, reportability is the only field that is changed via this process. The record's reportability is changed to auditable when the AFL status is set to "Not a Reportable Cancer".
- **AFUP Update** – Patient set data were updated when a response to Active Follow-up (AFUP) was entered.
- **Auto-Build Pat Set** – A single audit log entry is made when a patient set is auto-built from a record.
- **Auto-Coding** - A change was made to a record as it entered the workflow. Registry-specific coding algorithms are implemented in this automated workflow task.
- **Auto-Consolidation** - The change was made during an automated process to consolidate incoming data. Typically, this would involve the consolidation of follow-up data. In most registry configurations, the consolidation of other data is performed in manual tasks.
- **Auto-Screening** – The change was made during the automated screening task. This task sets the reportability status of an incoming record.
- **Consolidate** - The change was made by a user who edited the data during a Consolidate worklist task.
- **Consolidate FUP**- The change was made by a user who edited the data during a Consolidate FUP worklist task.
- **Data Entry** – A single audit log entry is made when the record is saved during a data entry session.

- **Direct Database Update** – Record or patient set data were modified by an external script (audit log entries are not created by all external scripts).
- **Edits Sys Task** – The change was made by a SEER*DMS "polisher" which executed after the Patient Set Edits system task completed. A polisher is a background utility that, according to SEER or registry-defined algorithms, imputes data values based on values in other fields.
- **Image Data Entry** – Record data were updated during a manual Image Data Entry task. An Image Data Entry task is created for each image record imported via the autoloader.
- **Importing**– A single audit log entry is made when the record is imported from a data file.
- **Initial Conversion** – The change was made during an automated process to load data from the legacy system.
- **Link AFL Source** – This event indicates that a record was auto-linked to a patient set. The record that initiated an AFL is auto-linked to a patient set when its AFL is closed by an abstract linked to the patient set.
- **Mass Change** – Changes were made as specified in a mass change import. The Import ID is listed with the event information.
- **Polishers Sys Task** – The change was made by a SEER*DMS "polisher". A polisher is a background utility that, according to SEER or registry-defined algorithms, imputes data values based on values in other fields.
- **Rapid Follow-up** – Patient set data were updated during an automated Rapid Follow-up worklist task.
- **Resolve Pat Set Errors** – The change was made by a user who edited the data during a Resolve Patient Set Errors worklist task.
- **Resolve Record Errors** - The change was made by a user who edited the data during a Resolve Record Errors worklist task.
- **Screening** - The change was made by a user who edited the data during a Screening worklist task.
- **Visual Edit Pat Set** – The change was made by a user who edited the data during a Visual Edit Patient Set worklist task.

Finding Patient Data in SEER*DMS

You may use the SEER*DMS Patient Lookup or toolbar Quick Search to search the database for data related to a specific patient, or to search for a specific record or patient set by ID.

Select **View > Patients** to access the Patient Lookup:



SEER*DMS uses registry-specific matching algorithms to compare the patient information specified in the Matching Criteria to patient sets and records in the database. The **Standard** search provides a way to find all available data related to a specific patient. The **Advanced** search and **ID Lookup** enable you to review a particular set of records or patient sets or to review a sample of data for research or quality control purposes.

If your system permissions allow you to view or edit patient data, the **Lookup** search box will be displayed in the SEER*DMS User Bar. This provides a shortcut to the Patient Lookup for searches by Patient Set ID, Record ID, AFL ID, Import ID, date of birth (MM-DD-YYYY), name ("last name" or "last name, first name"), and SSN (dashes must be included, otherwise the system will assume that a numeric field is a patient set or record ID).

Chapter 20: Searching for Records and Patients contains detailed instructions for using the Patient Lookup and Quick Search.

SEER*DMS Database and Data Warehouse

The Oracle database that is accessed via the production version of the SEER*DMS application is referred to as the "live database" or "production database". This database contains up-to-the-minute changes made by the registry's editing or coding staff.

The "data warehouse" is a read-only version of the live database which is created nightly. To reduce load on the production database, it is recommended that external applications query the data warehouse unless the query requires updates made since the previous night. The data warehouse contains all tables in the live database, the NAACCR mart, and materialized views which are optimized for extracts, reporting, and other analysis. Views are aggregated, sometimes summarized, copies of data that are specifically structured for dynamic queries and analytics.

The "test server" database contains the data accessed by the test server's version of SEER*DMS. This database is updated on an ad hoc basis by registry IT staff. A registry system administrator executes a reload script that loads a snapshot of the production database from the previous night (the data are loaded from the same snapshot as the data warehouse). If you log into the test server and save a change, that change will immediately be reflected in reports and SQL queries that use the Test Server Database. However, there is no data warehouse associated with the test server. Therefore, you cannot use the extracts within the test server's application (SEER*DMS extracts use data from the data warehouse which is only available for the production database). In addition, some internal SEER*DMS reports require the data warehouse and will not be available.

The SEER*DMS Web Portal (https://seer.cancer.gov/seerdms/portal) is the most comprehensive resource of documentation related to the SEER*DMS databases. The following materials are available online:

- *Description of each table in the database*
- *Mapping of database fields to the labels used in SEER*DMS*
- *Mapping of database fields to NAACCR items*
- *Diagrams showing table relationships*
- *NAACCR Mart and NAACCR Extract Documentation*